# A Performance and Cost Assessment of Machine Learning Interatomic Potentials

Yunxing Zuo,[†] Chi Chen,[†] Xiangguo Li,[†] Zhi Deng,[†] Yiming Chen,[†] Jörg Behler,[‡] Gábor Csányi,[¶] Alexander V. Shapeev,[§] Aidan P. Thompson,[‖] Mitchell A. Wood,[‖] and Shyue Ping Ong[*,†]

[†]*Department of NanoEngineering, University of California San Diego, 9500 Gilman Dr, Mail Code 0448, La Jolla, CA 92093-0448, United States*

[‡]*Universität Göttingen, Institut für Physikalische Chemie, Theoretische Chemie, Tammannstraße 6, 37077 Göttingen, Germany*

[¶]*Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom*

[§]*Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow, 143026, Russia*

[‖]*Center for Computing Research, Sandia National Laboratories, Albuquerque, New Mexico 87185, United States*

E-mail: ongsp@eng.ucsd.edu

**Abstract**

Machine learning of the quantitative relationship between local environment descriptors and the potential energy surface of a system of atoms has emerged as a new frontier in the development of interatomic potentials (IAPs). Here, we present a comprehensive evaluation of ML-IAPs based on four local environment descriptors — atom-centered symmetry functions (ACSF), smooth overlap of atomic positions

(SOAP), the Spectral Neighbor Analysis Potential (SNAP) bispectrum components, and moment tensors — using a diverse data set generated using high-throughput density functional theory (DFT) calculations. The data set comprising bcc (Li, Mo) and fcc (Cu, Ni) metals and diamond group IV semiconductors (Si, Ge) is chosen to span a range of crystal structures and bonding. All descriptors studied show excellent performance in predicting energies and forces far surpassing that of classical IAPs, as well as predicting properties such as elastic constants and phonon dispersion curves. We observe a general trade-off between accuracy and the degrees of freedom of each model, and consequently computational cost. We will discuss these trade-offs in the context of model selection for molecular dynamics and other applications.

# Introduction

A fundamental input for atomistic simulations of materials is a description of the potential energy surface (PES) as a function of atomic positions. While quantum mechanics-based descriptions, such as those based on Kohn-Sham density functional theory (DFT),[1,2] are accurate and transferable across chemistries, their high cost and poor scaling (typically $O(n_e^3)$ or higher, where $n_e$ is the number of electrons)[3–5] limits simulations to $\sim$ 1000 atoms and hundreds of picoseconds. Hence, large-scale and long-time simulations traditionally rely on interatomic potentials (IAPs), which to date are in most cases empirical parameterizations of the PES based on physical functional forms that depend only on the atomic degrees of freedom.[6–8] IAPs gain linear scaling with respect to the number of atoms at the cost of accuracy and transferability.

In recent years, a modern alternative has emerged in the form of machine-learned IAPs (ML-IAPs), where the PES is described as a function of local environment descriptors that are invariant to translation, rotation and permutation of homonuclear atoms.[9,10] Examples of such potentials include the high-dimensional neural network potential (NNP),[11,12] the Gaussian approximation potential (GAP),[10,13,14] the Spectral Neighbor Analysis Potential

(SNAP),[15–18] moment tensor potentials (MTP),[19–21] among others.[22–35] A typical approach to training such potentials involves the generation of a sufficiently large and diverse data set of atomic configurations with corresponding energies, forces and stresses from DFT calculations, which are then used in the training of the ML-IAP based on one or several target metrics, such as minimizing the mean absolute or squared errors in predicted energies, forces, stresses or derived properties (e.g. elastic constants). ML-IAPs have been shown to be a remarkable improvement over traditional IAPs, in general, achieving near-DFT accuracy in predicting energies and forces across diverse chemistries and atomic configurations. Despite the fact that recent benchmark efforts[36–38] have demonstrated the remarkable performance of ML-IAPs, a critical gap that remains is a rigorous assessment of the relative strengths and weaknesses of ML-IAPs across a standardized data set, similar to what has been done for classical IAPs.[39–41]

In this work, we present a comprehensive performance comparison of four major ML-IAPs — GAP, MTP, NNP and SNAP. The four IAPs were evaluated in terms of their accuracy in reproducing DFT energies and forces, as well as material properties such as the equations of state, lattice parameter and elastic constants. An attempt was also made to assess the training data requirements of each ML-IAP and the relative computational cost based on the best-available current implementations. To ensure a fair comparison, standardized DFT data sets of six elements (Li, Mo, Cu, Ni, Si and Ge) with the same training/test sampling and similar fitting approaches were used. The elements were chosen to span diverse chemistries and bonding, e.g., bcc and fcc metals, main group and transition metals, and group IV semiconductors.

# Methods

## Machine learning interatomic potentials

The four ML-IAPs investigated in this work have already been extensively discussed in previous works and reviews.[9–21,42–45] All ML-IAPs express the potential energy as a sum of atomic energies that are a function of the local environment around each atom, but differ in the descriptors for these local environments and the ML approach/functional expression used to map the descriptors to the potential energy. The detailed formalism of all four ML-IAPs are provided in the Supplementary Information. Here, only a concise summary of the key concepts and model parameters behind the ML-IAPs in chronological order of development is provided to aid the reader in following the remainder of this paper.

1. **High-dimensional neural network potential (NNP).** The NNP uses atom-centered symmetry functions (ACSF)[46] to represent the atomic local environments and fully connected neural networks to describe the PES with respect to symmetry functions.[11,12] A separate neural network is used for each atom and each neural network processes the symmetry functions from local environments of the corresponding atom and outputs its atomic energy. The architecture of the neural network is defined by the number of hidden layers and the nodes in each layer, while the descriptor space is given by the following symmetry functions:

$$
\begin{aligned}
G_i^{\mathrm{atom,rad}} &= \sum_{j\neq i}^{N_{\mathrm{atom}}} e^{-\eta(R_{ij}-R_s)^2}\cdot f_c(R_{ij}),\\
G_i^{\mathrm{atom,ang}} &= 2^{1-\zeta}\sum_{j,k\neq i}^{N_{\mathrm{atom}}} (1+\lambda\cos\theta_{ijk})^{\zeta}\cdot e^{-\eta'(R_{ij}^2+R_{ik}^2+R_{jk}^2)}\cdot f_c(R_{ij})\cdot f_c(R_{ik})\cdot f_c(R_{jk}),
\end{aligned}
$$

where $R_{ij}$ is the distance between atom $i$ and neighbor atom $j$, $\eta$ is the width of the Gaussian and $R_s$ is the position shift over all neighboring atoms within the cutoff radius $R_c$, $\eta'$ is the width of the Gaussian basis and $\zeta$ controls the angular resolution.

$f_c(R_{ij})$ is a cutoff function, defined as follows:

$$f_c(R_{ij}) = \begin{cases} 0.5 \cdot [\cos(\frac{\pi R_{ij}}{R_c}) + 1], & \text{for } R_{ij} \leq R_c \\ 0.0, & \text{for } R_{ij} > R_c. \end{cases} \tag{1}$$

These hyperparameters were optimized to minimize the root-mean- square errors of energies and forces for each chemistry. The NNP model has shown great performance for Si,[11] $TiO_2$,[47] water[48] and solid-liquid interfaces,[49] metal-organic frameworks,[50] and has been extended to incorporate long-range electrostatics for ionic systems such as ZnO[51] and $Li_3PO_4$.[52]

2. **Gaussian Approximation Potential (GAP).** The GAP calculates the similarity between atomic configurations based on a smooth-overlap of atomic positions (SOAP)[10,53] kernel, which is then used in a Gaussian process model. In SOAP, the Gaussian-smeared atomic neighbor densities $\rho_i(\boldsymbol{R})$ are expanded in spherical harmonics as follows:

$$\rho_i(\boldsymbol{R}) = \sum_j f_c(R_{ij}) \cdot \exp(-\frac{|\boldsymbol{R} - \boldsymbol{R}_{ij}|^2}{2\sigma^2_{\text{atom}}}) = \sum_{nlm} c_{nlm}\ g_n(R) Y_{lm}(\hat{\boldsymbol{R}}), \tag{2}$$

The spherical power spectrum vector, which is in turn the square of expansion coefficients,

$$p_{n_1 n_2 l}(\boldsymbol{R}_i) = \sum_{m=-l}^{l} c^*_{n_1 lm} c_{n_2 lm}, \tag{3}$$

can be used to construct the SOAP kernel while raised to a positive integer power $\zeta$ to accentuate the sensitivity of the kernel,[10]

$$\boldsymbol{K}(\boldsymbol{R}, \boldsymbol{R'}) = \sum_{n_1 n_2 l} (p_{n_1 n_2 l}(\boldsymbol{R}) p_{n_1 n_2 l}(\boldsymbol{R'}))^{\zeta}, \tag{4}$$

In the above equations, $\sigma_{\text{atom}}$ is a smoothness controlling the Gaussian smearing, and $n_{\max}$ and $l_{\max}$ determine the maximum powers for radial components and angular components in spherical harmonics expansion, respectively.[10] These hyperparameters, as well as the number of reference atomic configurations used in Gaussian process, are optimized in the fitting procedure to obtain optimal performance. The GAP has been developed for transition metals,[13,14] main group elements,[54–56] diamond semiconductors[57,58] as well as multi-component systems.[44]

3. **Spectral Neighbor Analysis Potential (SNAP).** The SNAP uses the coefficients of the bispectrum of the atomic neighbor density functions[10] as descriptors. In the original formulation of SNAP, a linear model between energies and bispectrum components is assumed.[15] Recently, a quadratic model (denoted as qSNAP in this work)[59] has been developed, which extends the linear SNAP energy model to include all distinct pairwise products of bispectrum components. In this work, both linear and quadratic SNAP models were investigated. The critical hyperparameters influencing model performance are the cutoff radius and $J_{\max}$, which limits the indices $j_1$, $j_2$, $j$ in Clebsch-Gordan coupling coefficients $H^{jmm\prime}_{\substack{j_1 m_1 m_1' \\ j_2 m_2 m_2'}}$ in construction of the bispectrum components:

$$B_{j_1,j_2,j} = \sum_{m_1,m_1'=-j_1}^{j_1} \sum_{m_2,m_2'=-j_2}^{j_2} \sum_{m,m'=-j}^{j} (u^j_{m,m'})^* \times H^{jmm\prime}_{\substack{j_1 m_1 m_1' \\ j_2 m_2 m_2'}} u^{j_1}_{m_1,m_1'} u^{j_2}_{m_2,m_2'}, \tag{5}$$

where $u^j_{m,m'}$ are coefficients in 4-dimensional hyper-spherical harmonics expansion of neighbor density function:

$$\rho_i(\mathbf{R}) = \sum_{j=0}^{\infty} \sum_{m,m'=-j}^{j} u^j_{m,m'} U^j_{m,m'}, \tag{6}$$

The SNAP model as well as qSNAP model has demonstrated great success in transition

metals[15–17,59] as well as binary systems.[17,18,45]

4. **Moment Tensor Potential (MTP)**. The MTP[19] devises rotationally-covariant tensors

$$M_{\mu,\nu}(\boldsymbol{R}) = \sum_j f_\mu(R_{ij}) \underbrace{\boldsymbol{R}_{ij} \otimes \cdots \otimes \boldsymbol{R}_{ij}}_{\nu \text{ times}}, \tag{7}$$

to describe the atomic local environments. Here $f_\mu$ are the radial functions, and $\boldsymbol{R}_{ij} \otimes \cdots \otimes \boldsymbol{R}_{ij}$ are tensors of rank $\nu$ encoding angular information about the atomic environment. The rank $\nu$ can be large enough to approximate any arbitrary interactions. MTP then contracts these tensors to a scalar that yields rotationally-invariant basis functions, and applies linear regression to correlate the energies with the basis functions. The performance of MTP is controlled by the polynomial power-like metric, which defines what tensors and how many times are contracted. The MTP model has been successfully applied to metals,[19,20,60] boron,[61] binary and ternary alloys[21] as well as gas-phase chemical reactions.[62]

## DFT Data Sets

A comprehensive DFT data set was generated for six elements - Li, Mo, Ni, Cu, Si and Ge. These elements were chosen to span a variety of chemistries (main group metal, transition metal and semiconductor), crystal structures (bcc, fcc, and diamond) and bonding types (metallic and covalent). For each element, we generated a set of structures with diverse coverage of atomic local environment space, as follows:

(1) The ground-state crystal for each element.

(2) Strained structures constructed by applying strains of $-10\%$ to 10% at 2% intervals to the bulk supercell in six different modes, as described in the work by de Jong et al.[63]. The supercells used are the $3 \times 3 \times 3$, $3 \times 3 \times 3$ and $2 \times 2 \times 2$ of the conventional bcc, fcc and diamond unit cells, respectively.

(3) Slab structures up to a maximum Miller index of three, including (100), (110), (111), (210), (211), (310), (311), (320), (321), (322), (331), and (332), as obtained from the Crystalium database.[64,65]

(4) *NVT ab initio* molecular dynamics (AIMD) simulations of the bulk supercells (similar to those in (2)) performed at 300 K and 0.5×, 0.9×, 1.5×, 2.0× of the melting point of each element with a time step of 2 fs. The bulk supercells were heated from 0 K to the target temperatures and equilibrated for 20,000 time steps. A total of 20 snapshots were obtained from the subsequent production run in each AIMD simulation at an interval of 0.1 ps. The radial distribution function (RDF) analysis confirms that the structures obtained from AIMD simulations above melting point are liquid-like in the atomic environments (see Figure S1).

(5) *NVT* AIMD simulations of the bulk supercells (similar to those in (2)) with a single vacancy performed at 300 K and 2.0× of the melting point of each element. The bulk supercells were heated from 0 K to the target temperatures and equilibrated for 20,000 time steps. A total of 40 snapshots were obtained from the subsequent production run of each AIMD simulation at an interval of 0.1 ps.

All DFT calculations were carried out using the Vienna *ab initio* simulation package (VASP)[66] version 5.4.1 within the projector augmented wave approach.[67] The Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA)[68] was adopted for the exchange-correlation functional. The kinetic-energy cutoff was set to 520 eV and the $k$-point mesh was $4 \times 4 \times 4$ for the Mo, Ni, Cu, Si and Ge supercells, and $3 \times 3 \times 3$ for the Li supercells. The electronic energy and atomic force components were converged to within $10^{-5}$ eV and 0.02 eV/Å, respectively, in line with previous works.[16,17] The AIMD simulations were carried out with a single $\Gamma$ $k$ point and were non-spin-polarized, but static calculations using the same parameters as the rest of the data were carried out on the snapshots to obtain consistent energies and forces. All structure manipulations and analyses of DFT compu-

tations were carried out using Python Materials Genomics (Pymatgen)[69] library, and the automation of calculations was performed using the Fireworks software.[70]

## Optimization scheme

Figure 1 provides an overview of the general data generation and potential development scheme. The training data set was first generated via DFT static calculations on the four categories of structures. The optimization procedure comprised two loops. In the inner loop, sampled structures in the database were transformed into atomic descriptors (e.g., bispectrum components for SNAP and symmetry functions for NNP), which were then fed into the corresponding ML model together with the DFT energies, forces, and stresses as the targets of training. The data was apportioned into training and test sets with a 90:10 split. The parameters of the ML models were optimized during the training process. In the outer loop, the ML model trained in the inner loop was used to predict basic material properties (e.g., elastic tensors), and the differences between the predicted and reference values were then used to determine the optimal hyperparameters for each ML-IAPs. In this work, we adopted a combination of the grid search algorithm and differential evolution algorithm to perform hyperparameters optimization for different ML-IAPs.

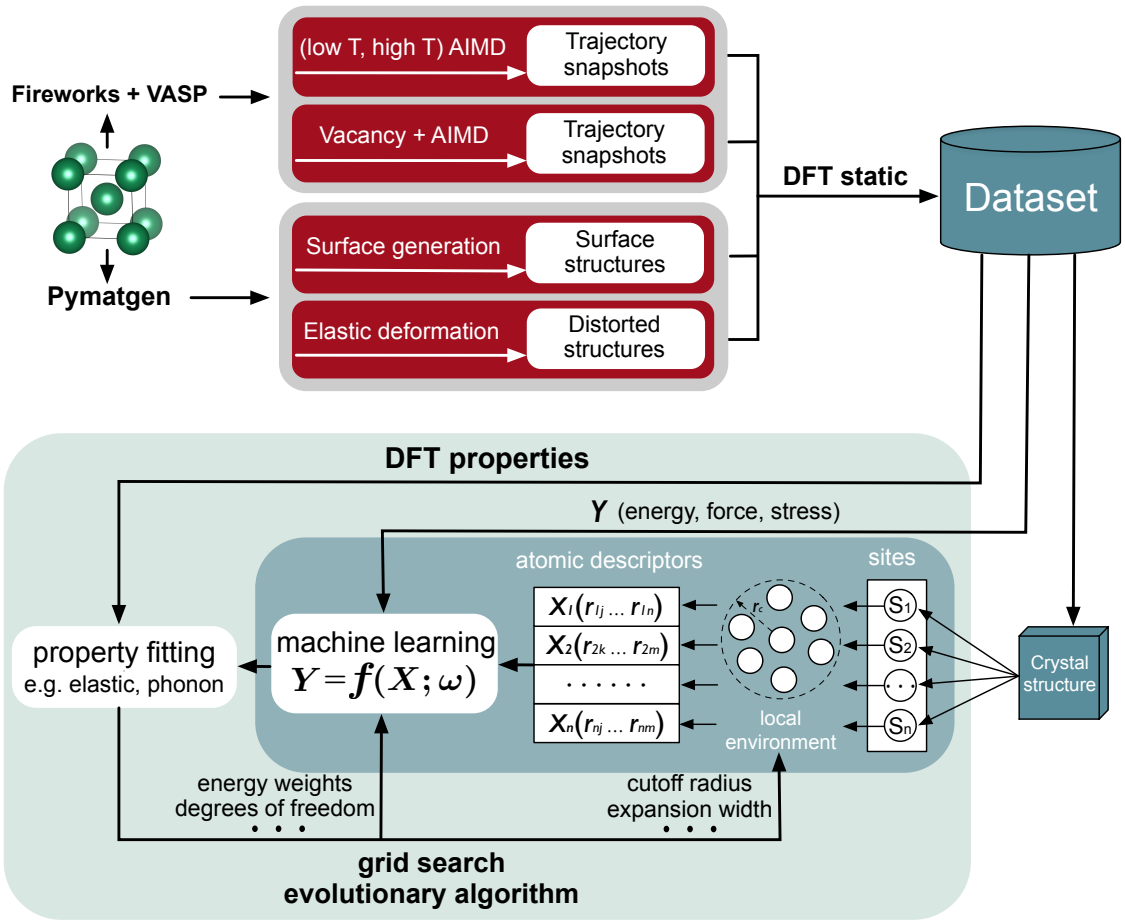



Figure 1: Machine-learning interatomic potential development workflow.

## Data and code availability

To facilitate the reuse and reproduction of our results, the code, data and optimized ML models in this work are published open-source on Github (`https://github.com/materialsvirtuallab/mlearn`). The code includes high-level Python interfaces for ML-IAPs development as well as LAMMPS material properties calculators.

# Results

## Optimized model parameters

The optimized coefficients and hyperparameters for each ML-IAP are reported in Supplementary Information (see Table S1 - S11). Here, we will limit our discussions to a parameter that is common to all ML-IAPs - the cutoff radius - and present a convergence study of each ML-IAP with the number of degrees of freedom of the model.

The cutoff radius determines the maximum range of interatomic interactions, and hence, has a critical effect on the prediction performance of ML-IAPs. Table 1 provides the optimized cutoff radii of different ML-IAPs across different chemistries. Different ML-IAPs yield similar optimized cutoff radii for the same elemental system. The optimized cutoff radii are between the second nearest neighbor (2NN) and 3NN distance for fcc elements (Cu, Ni), between 3NN and 4NN distances for the bcc (Li, Mo) and diamond (Ge and Si) elements. These observations are consistent with those from previous traditional and ML IAP development efforts, where typically 2NN interactions are found to suffice for fcc metals,[71,72] while contributions from 3NN cannot be ignored for bcc metals[13,19,20,73,74] and diamond systems.[75,76]

Table 1: Optimized cutoff radius for each element for each ML-IAP.

| | **fcc** | | **bcc** | | **diamond** | |
|---|---|---|---|---|---|---|
| cutoff radius (Å) | Ni | Cu | Li | Mo | Si | Ge |
| GAP | 3.9 | 3.9 | 4.8 | 5.2 | 5.4 | 5.4 |
| MTP | 4.0 | 3.9 | 5.1 | 5.2 | 4.7 | 5.1 |
| NNP | 3.9 | 4.1 | 5.2 | 5.2 | 5.5 | 5.6 |
| SNAP | 4.1 | 4.1 | 5.1 | 4.6 | 4.9 | 5.5 |
| qSNAP | 3.8 | 3.9 | 5.1 | 5.2 | 4.8 | 4.9 |

The number of degrees of freedom (DOF), e.g., the number of weights and biases for the NNP and number of representative points in GAP, has a strong effect on the accuracy

and computational cost of each ML-IAP. Figure 2 illustrates the trade-off between computational cost and test error under varying DOFs for each fitted Mo ML-IAP. Similar results are obtained for other systems (see Figure S5 - S10). It should be noted that the relative computational costs are based on the most efficient available implementations[10,15,19,59,77] of each ML-IAP at this time in LAMMPS[78] and performed on a single CPU core of Intel i7-6850k 3.6 GHz with $18 \times 18 \times 18$ bulk supercell containing 11,664 atoms for Mo system. Future implementations and optimizations, such as to the evaluation of the local environment descriptor, may improve on these results, as discussed in a recent work.[79] A dashed Pareto frontier is drawn in Figure 2a to represent points at which better accuracy can only be attained at the price of greater computational cost[80] and the black arrows indicate "optimal" configurations for each model in terms of the trade-off between test error and computational cost. These "optimal" configurations were used for subsequent accuracy comparisons in energies, forces and properties. We find that the "optimal" MTP, NNP, SNAP and qSNAP models tend to be two orders of magnitude less computationally expensive than the "optimal" GAP model. The MTP models generally lie close to the Pareto frontier, exhibiting an excellent balance between model accuracy and computational efficiency. For the SNAP and qSNAP models, the descriptor space (i.e., bispectrum components) is determined by the parameter $J_{\max}$. We find that the rate-limiting step is the calculation of bispectrum and the computation of quadratic terms in qSNAP has only a marginal effect on the computational cost.[59] However, we find that the substantial expansion in the number of DOF in the qSNAP model leads to over-fitting for larger values of $J_{\max}$ (see Figure 2b), which can be attributed to the large number of fitted coefficients of its formalism. For the GAP model, the computational cost is linearly related to the number of kernels used in Gaussian process regression.[13] It should be noted that classical IAPs remain substantially lower in computational cost than ML-IAPs by 2-3 orders of magnitude. The computational time for prediction using all ML-IAPs and equivalent classical IAPs are given in Table S12.

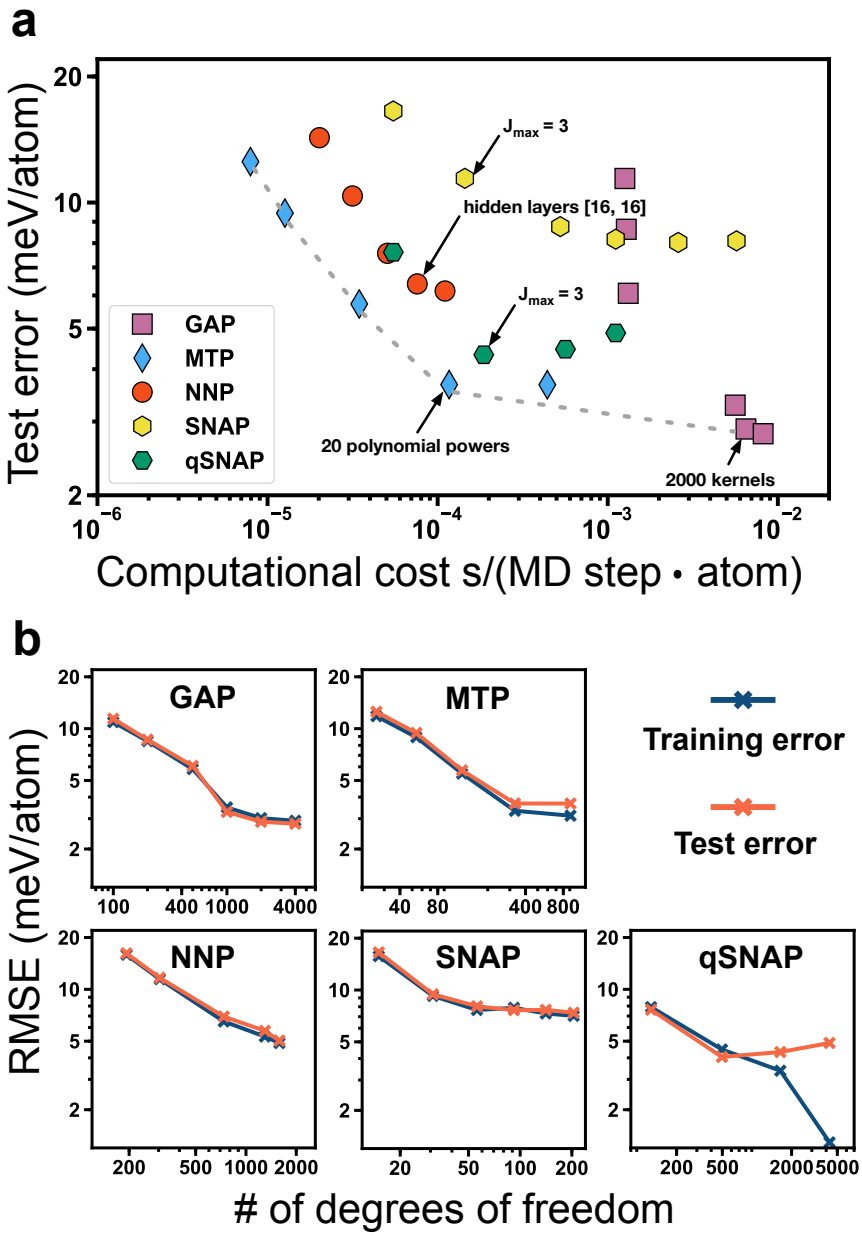



Figure 2: (a) Test error versus computational cost for the Mo system. The grey dashed line indicates an approximate Pareto frontier formed by the convex hull of points lying on the bottom left of the chart. This Pareto frontier represents an optimal trade-off between accuracy and computational cost. Timings were performed by LAMMPS calculations on a single CPU core of Intel i7-6850k 3.6 GHz. Black arrows denote the "optimal" configuration for each ML-IAP that was used in subsequent comparisons. (b) Plots of the training and test errors versus the number of degrees of freedom for each ML-IAP.

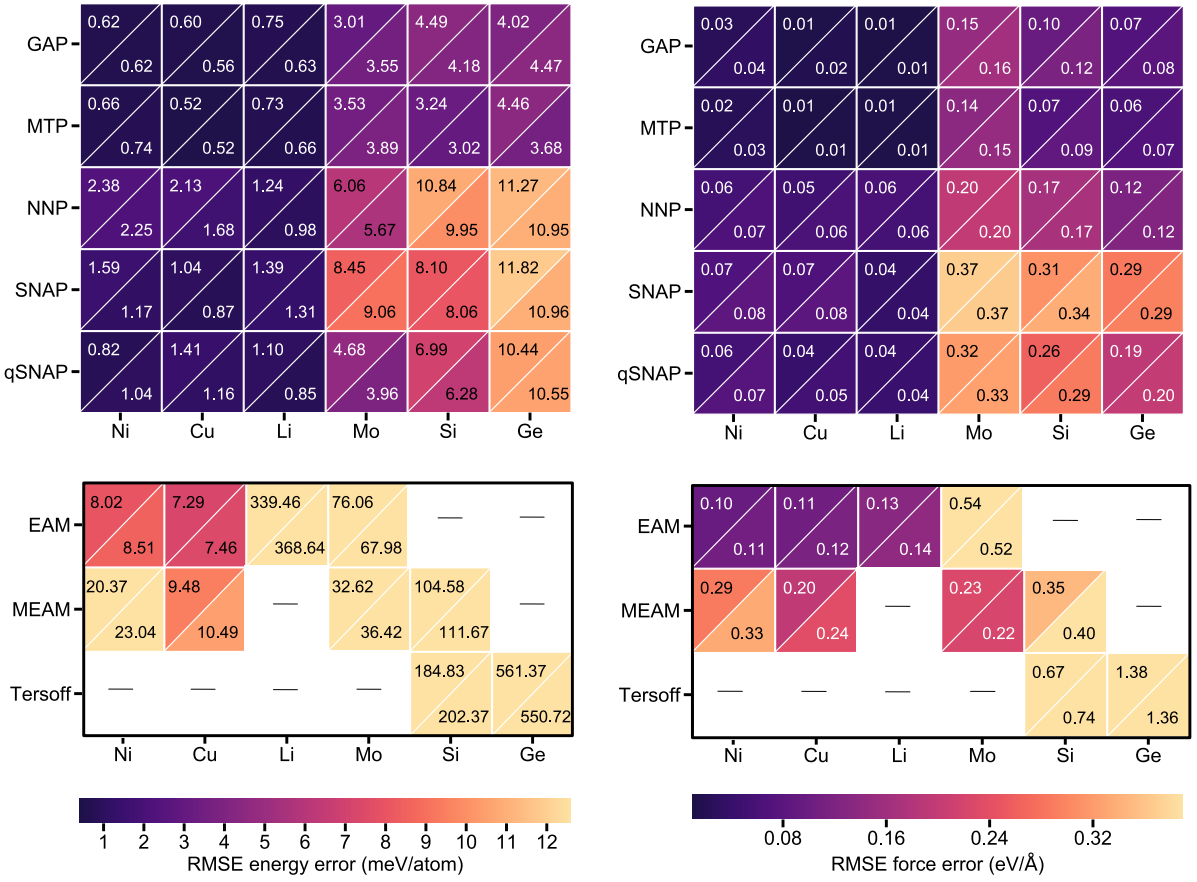
## Accuracy in energies and forces

Figure 3 provides a comparison of the root-mean-square errors (RMSEs) in energies and forces for the four ML-IAPs and best-available classical IAPs relative to DFT. All ML-IAPs show extremely good performance across all elements studied, achieving RMSEs in energies and forces that are far lower than best-available traditional IAPs for each element. It should be noted that differences in RMSEs between ML-IAPs are on the scale of meV atom$^{-1}$ in energies and 0.1 eV Å$^{-1}$ in forces; hence, any subsequent discussion on the relative performances of the ML-IAPs should be viewed in the context that even the largest differences in accuracy between the ML-IAPs are already close to the limits of DFT error. In all cases, the training and test errors are similar, indicating no over-fitting for the optimized ML-IAPs. The energy and force errors of ML-IAPs are marginally larger than those reported in a recent work[81] on Cu and Ge. However, this is likely due to the fact that the data in this work contains a diverse range of structures sampled from AIMD simulations at a much larger time intervals (0.1 ps vs 10 fs in the previous work) as well as strained structures, surface structures and vacancy-containing structures.

The GAP and MTP models generally have the lowest RMSEs in energies and forces. The highest RMSEs in energies are observed for the SNAP models and NNP models. It is well-known that neural network-based models often require larger data sets for best performance; previous NNP models have been trained on thousands or tens of thousands of structures,[36,82] while only hundreds of structures are used in training the current ML-IAPs. Nevertheless, the NNP models still show surprisingly good performance for bcc systems. The qSNAP models' performances are between those of the GAP and NNP. In general, the qSNAP models have moderately lower RMSEs than the linear SNAP, though at the expense of a large expansion in the number of parameters.

In terms of chemistries, we find that the lowest RMSEs in energies are observed for the fcc systems, followed by the bcc systems, and the highest RMSEs are observed for the diamond systems. Very low RMSEs in forces are observed across all ML-IAPs for Cu, Ni

and Li, while relatively higher RMSEs in forces are observed for Mo, a metal with higher modulus and larger force distributions. Higher RMSEs in forces are also observed for the diamond semiconductors. These trends are generally consistent across all ML-IAPs studied. The RMSEs normalized to the ground state energy per atom for each element is given in Figure S2) further supports the better accuracy for the metals (both fcc and bcc) relative to the diamond systems, and the much better performance of the ML-IAPs over the classical IAPs.




(a) Root-mean-square errors in predicted energies (b) Root-mean-square errors in predicted forces

Figure 3: Root-mean-square errors in (a) predicted energies (b) predicted forces for all four ML-IAPs as well as traditional IAPs (EAM,[83,84] MEAM,[85–87] Tersoff[88,89]). The upper left and lower right triangles within each cell represent training and test errors, respectively.

We have also performed a study of the convergence of the ML-IAPs with training data size using Mo as the benchmark system given that it is a bcc metal (for which traditional

IAPs tend to perform poorly) with large force distributions. Here, the length of the AIMD simulations was increased four-fold, and more training structures were sampled at the same time interval. The convergence results are shown in Figure 4. While the prediction errors of all models decrease with increase in the number of training structures, the most substantial improvements in accuracy, especially in predicted energies, are observed for the NNP and qSNAP models. The SNAP Mo model appears to have converged in energy and force at a training data size of $\sim$ 400 structures, respectively. For the qSNAP, additional training structures offer modest improvements in force accuracy, but large improvements in energy accuracy, which implies qSNAP has a higher "learning rate" in terms of energy accuracy and a lower "learning rate" in terms of force accuracy than NNP. This can also be seen in Figure S3, which showed the log-log plot of accuracies versus the size of training data and the slope of the line as the "learning capability" of different ML-IAPs. Indeed, it is possible that the NNP and qSNAP Mo models have not been converged with respect to accuracy in energies even at $\sim$ 800 training structures. We have not attempted to further converge these models in view of the computational expense involved.

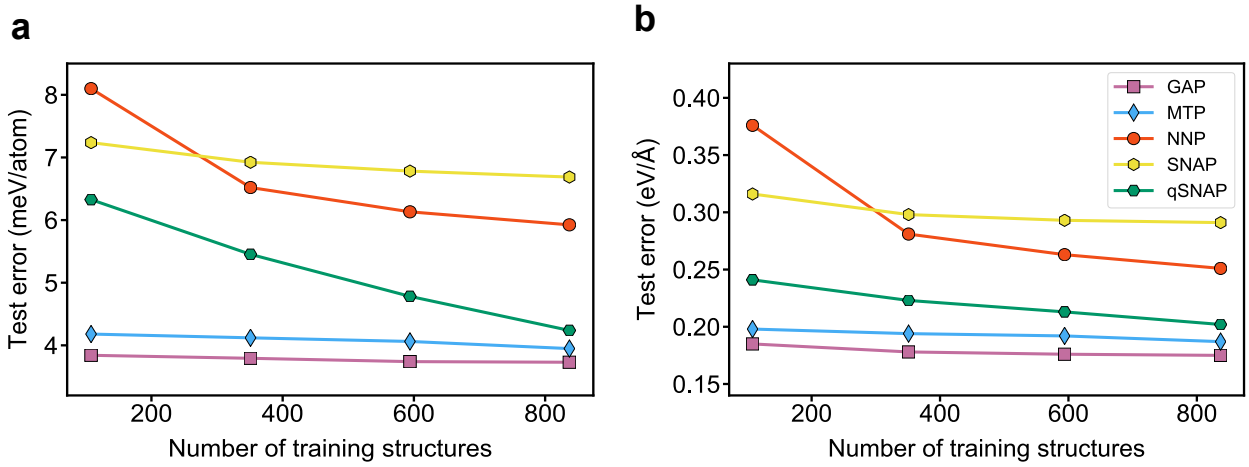



Figure 4: RMSEs in predicted (a) energies (b) forces of the test set versus the size of the training data for the ML-IAP Mo models.

## Accuracy in material properties

The accuracy in predicting basic material properties is critical for evaluating the performance of ML-IAPs. Here, we perform the climbing-image nudged elastic band (CI-NEB) method[90] as well as molecular dynamics (MD) with ML-IAPs to obtain the cubic lattice parameter, elastic constants, migration energies and vacancy formation energies. The comparison of these predicted material properties with respect to the DFT values is provided in Table 2. The performances of all ML-IAPs are generally excellent, with lattice parameters within 0.1-2.0% of the DFT values and elastic constants that are typically within 10% of DFT values. It should be noted that the large percentage error in Li for elastic constants is due to the small reference values. The MTP, SNAP and qSNAP models perform well on elastic constants on fcc and bcc systems, but exhibit slightly higher prediction errors in the diamond systems. A possible explanation for the slightly poorer prediction of elastic constants of the NNP model could be the limitation of the size of training data, which restricts the potential of a fully connected neural network. However, it should be noted that despite the slightly higher prediction errors of elastic components for the NNP model, its prediction errors of Voigt-Reuss-Hill approximated bulk modulus[91] across various elemental systems are in good agreement with DFT reference values.

In terms of diffusion properties, the GAP and MTP models perform well across different chemistries, with most of the prediction errors within 10% of DFT values, albeit with a moderate underestimate of the migration energy for diamond systems, in line with the previous study.[57] While SNAP and qSNAP models show high accuracy in predicting diffusion properties for fcc systems, they considerably underestimate the vacancy formation energy as well as activation barrier for diamond systems. It is noteworthy that all ML-IAPs overestimate the migration energy of Mo system by more than 20%, which has also been observed in a previous work.[16]

Table 2: Calculated cubic lattice parameter $a$, elastic constants ($c_{ij}$), Voigt-Reuss-Hill bulk modulus $B_{\mathrm{VRH}}$, migration energy ($E_m$), vacancy formation energy ($E_v$) as well as activation barrier for vacancy diffusion ($E_a = E_v + E_m$) with DFT and the four ML-IAPs. Lowest absolute errors with respect to DFT for each property are bolded for ease of reference. Error percentages with respect to DFT values are shown in parentheses.

| | **DFT** | **GAP** | **MTP** | **NNP** | **SNAP** | **qSNAP** |
|---|---|---|---|---|---|---|
| **Ni** | | | | | | |
| $a$ (Å) | 3.508 | 3.523 (0.4%) | 3.522 (0.4%) | 3.523 (0.4%) | 3.522 (0.4%) | **3.521 (0.4%)** |
| $c_{11}$ (GPa) | 276 | 281 (1.8%) | 284 (2.9%) | **274 (-0.8%)** | 283 (2.5%) | 267 (−3.3%) |
| $c_{12}$ (GPa) | 159 | **159 (0.0%)** | 172 (8.2%) | 169 (6.3%) | 168 (5.7%) | 155 (−2.5%) |
| $c_{44}$ (GPa) | 132 | 126 (−4.5%) | 127 (-3.8%) | 113 (−14.4%) | **129 (-2.3%)** | 125 (−5.3%) |
| $B_{\mathrm{VRH}}$ (GPa) | 198 | **200 (1.0%)** | 209 (5.6%) | 204 (3.0%) | 206 (4.0%) | 193 (−2.5%) |
| $E_v$ (eV) | 1.49 | 1.46 (−2.0%) | 1.43 (−4.0%) | 1.65 (10.7%) | **1.47 (-1.3%)** | **1.47 (-1.3%)** |
| $E_m$ (eV) | 1.12 | 1.14 (1.8%) | 1.11 (−0.9%) | 1.14 (1.8%) | **1.12 (0.0%)** | 1.05 (−6.3%) |
| $E_a$ (eV) | 2.61 | **2.60 (-0.4%)** | 2.54 (−2.7%) | 2.79 (6.9%) | 2.59 (−0.8%) | 2.52 (−3.4%) |
| **Cu** | | | | | | |
| $a$ (Å) | 3.621 | **3.634 (0.4%)** | 3.636 (0.4%) | 3.637 (0.4%) | 3.634 (0.4%) | **3.636 (0.4%)** |
| $c_{11}$ (GPa) | 173 | **175 (1.2%)** | 177 (2.3%) | 182 (5.2%) | 178 (2.9%) | 178 (2.9%) |
| $c_{12}$ (GPa) | 133 | 120 (−9.8%) | 120 (9.8%) | 125 (−6.0%) | **126 (**−5.3%) | 124 (−6.8%) |
| $c_{44}$ (GPa) | 88 | 82 (−6.8%) | 81 (−8.0%) | 76 (−13.6%) | **86 (-2.3%)** | 82 (−6.8%) |
| $B_{\mathrm{VRH}}$ (GPa) | 146 | 138 (−5.5%) | 139 (−4.8%) | **144 (-1.4%)** | 143 (−2.1%) | 142 (−2.7%) |
| $E_v$ (eV) | 1.15 | 1.05 (−8.7%) | 1.10 (−4.3%) | 1.23 (7.0%) | 1.19 (3.5%) | **1.15 (0.0%)** |
| $E_m$ (eV) | 0.79 | 0.76 (−3.8%) | **0.77 (-2.5%)** | **0.77 (-2.5%)** | 0.82 (3.8%) | 0.74 (−6.3%) |
| $E_a$ (eV) | 1.94 | 1.81 (−6.7%) | 1.87 (−3.6%) | 2.00 (3.1%) | 2.01 (3.6%) | **1.89 (-2.6%)** |
| **Li** | | | | | | |
| $a$ (Å) | 3.427 | 3.450 (0.7%) | 3.446 (0.6%) | **3.434 (0.2%)** | 3.506 (2.3%) | 3.469 (1.2%) |
| $c_{11}$ (GPa) | 15 | 18 (20.0%) | **14 (-6.7%)** | 17 (13.3%) | 18 (20.0%) | 12 (−20.0%) |
| $c_{12}$ (GPa) | 13 | 14 (7.7%) | **13 (0.0%)** | 12 (−7.7%) | 7 (−46.2%) | 6 (−53.8%) |
| $c_{44}$ (GPa) | 11 | 12 (9.1%) | **11 (0.0%)** | 12 (9.1%) | 10 (−9.1%) | **11 (0.0%)** |
| $B_{\mathrm{VRH}}$ (GPa) | 14 | **15 (7.1%)** | **13 (-7.1%)** | **13 (-7.1%)** | 11 (−21.4%) | 8 (−42.9%) |
| $E_v$ (eV) | 0.62 | 0.56 (−9.7%) | 0.53 (−14.5%) | 0.50 (−19.4%) | **0.63 (1.6%)** | 0.58 (−6.5%) |
| $E_m$ (eV) | 0.06 | **0.06 (0.0%)** | 0.08 (33.3%) | 0.05 (−16.7%) | 0.09 (50.0%) | 0.09 (50.0%) |
| $E_a$ (eV) | 0.68 | 0.62 (−8.8%) | 0.61 (−10.3%) | 0.55 (−19.1%) | 0.72 (5.9%) | **0.67 (-1.5%)** |

Continued on next page

**Table 2 – continued from previous page**

| | DFT | GAP | MTP | NNP | SNAP | qSNAP |
|---|---|---|---|---|---|---|
| **Mo** | | | | | | |
| $a$ (Å) | 3.168 | **3.168 (0.0%)** | 3.169 (0.0%) | 3.165 (−0.1%) | 3.169 (0.0%) | 3.170 (0.1%) |
| $c_{11}$ (GPa) | 472 | 481 (1.9%) | **472 (0.0%)** | 441 (−6.6%) | 457 (−3.2%) | 436 (−7.6%) |
| $c_{12}$ (GPa) | 158 | 169 (7.0%) | 154 (−2.5%) | 192 (21.5%) | **158 (0.0%)** | 166 (5.1%) |
| $c_{44}$ (GPa) | 106 | 112 (5.7%) | 103 (−2.8%) | 114 (7.5%) | 109 (2.8%) | **104 (-1.9%)** |
| $B_{\mathrm{VRH}}$ (GPa) | 263 | 271 (3.8%) | **260 (-1.1%)** | **266 (1.1%)** | 258 (−1.9%) | 256 (−2.7%) |
| $E_v$ (eV) | 2.70 | **2.68 (-0.7%)** | 2.61 (−3.3%) | 2.94 (8.9%) | **2.72 (0.7%)** | 2.79 (3.3%) |
| $E_m$ (eV) | 1.22 | 1.60 (31.1%) | 1.51 (23.8%) | 1.59 (30.3%) | **1.49 (22.1%)** | 1.50 (23.0%) |
| $E_a$ (eV) | 3.92 | 4.28 (9.2%) | **4.12 (5.1%)** | 4.53 (15.6%) | 4.21 (7.4%) | 4.29 (9.4%) |
| **Si** | | | | | | |
| $a$ (Å) | 5.469 | 5.458 (−0.2%) | **5.465 (-0.1%)** | 5.501 (0.6%) | 5.466 (0.1%) | 5.464 (-0.1%) |
| $c_{11}$ (GPa) | 156 | 168 (7.7%) | **155 (-0.6%)** | 141 (−9.6%) | 128 (−17.9%) | **155 (-0.6%)** |
| $c_{12}$ (GPa) | 65 | **62 (-4.6%)** | 76 (16.9%) | **62 (-4.6%)** | 75 (15.4%) | 58 (−10.8%) |
| $c_{44}$ (GPa) | 76 | 69 (−9.2%) | **75 (-1.3%)** | 55 (−27.6%) | 71 (−6.6%) | 69 (−9.2%) |
| $B_{\mathrm{VRH}}$ (GPa) | 95 | **97 (2.1%)** | 102 (7.4%) | 89 (−6.3%) | **93 (-2.1%)** | 90 (−5.3%) |
| $E_v$ (eV) | 3.25 | 3.04 (−6.5%) | **3.11 (-4.3%)** | 2.60 (−20.0%) | 2.71 (−16.6%) | 2.37 (−27.1%) |
| $E_m$ (eV) | 0.21 | **0.21 (0.0%)** | 0.16 (−23.8%) | **0.21 (0.0%)** | 0.26 (23.8%) | 0.20 (−4.7%) |
| $E_a$ (eV) | 3.46 | 3.25 (−6.1%) | **3.27 (-5.5%)** | 2.81 (−18.8%) | 2.97 (−14.2%) | 2.57 (−25.7%) |
| **Ge** | | | | | | |
| $a$ (Å) | 5.763 | 5.777 (0.2%) | **5.770 (0.1%)** | 5.789 (0.5%) | 5.773 (0.2%) | 5.775 (0.2%) |
| $c_{11}$ (GPa) | 116 | 127 (9.5%) | 106 (−8.6%) | 98 (−15.5%) | 101 (−12.9%) | **121 (4.3%)** |
| $c_{12}$ (GPa) | 48 | **45 (-6.3%)** | 54 (12.5%) | 54 (12.5%) | 41 (−14.6%) | 43 (−10.4%) |
| $c_{44}$ (GPa) | 58 | 54 (−6.9%) | **55 (-5.2%)** | 43 (−25.9%) | 54 (−6.9%) | 50 (−13.8%) |
| $B_{\mathrm{VRH}}$ (GPa) | 71 | 72 (1.4%) | **71 (0.0%)** | 69(−2.8%) | 61 (−14.1%) | 69 (−2.8%) |
| $E_v$ (eV) | 2.19 | **2.10 (-4.1%)** | 1.98 (−9.6%) | 1.97 (−10.0%) | 1.77 (−19.2%) | 1.67 (−23.7%) |
| $E_m$ (eV) | 0.19 | 0.17 (−10.5%) | 0.17 (−10.5%) | **0.20 (5.3%)** | 0.28 (47.4%) | **0.18 (-5.3%)** |
| $E_a$ (eV) | 2.38 | **2.27 (-4.6%)** | 2.15 (−9.7%) | 2.18 (−8.8%) | 2.05 (−13.9) | 1.85 (−22.3%) |

We have also calculated the relaxed (011) $\gamma$ surface for the generalized stacking faults

(GSFs) along the $[\bar{1}\bar{1}1]$ direction[86,92] for Mo, (111) glide plane along the [112] direction[57,93] for Si, and (111) $\gamma$ surface along the [112] direction[94,95] for Ni and Cu, as these have been already been studied extensively in previous works.[57,86,92–95] From Figure 5 and Table 3, it can be observed that all ML-IAPs are able to reproduce the major qualitative features of the relaxed (011) $\gamma$ surface and the correct trend in the unstable stacking fault energy for all four systems. The MTP models generally yield $\gamma_{us}$ that are closest to DFT for Si (111)[112], Ni(111)[112] and Cu(111)[112]. All ML-IAPs significantly underestimate the Mo(011)$[\bar{1}\bar{1}1]$ $\gamma$ surface, which is consistent with previous ML-IAP studies of bcc metals.[15] The NNP models exhibit the largest deviation from DFT for Mo(011)[112], Ni(111)[112] and Cu(111)[112] $\gamma$ surfaces. In particular, the NNP models predict a symmetric $\gamma$ surface with a near-zero intrinsic stacking fault energy for the fcc metals Ni and Cu.

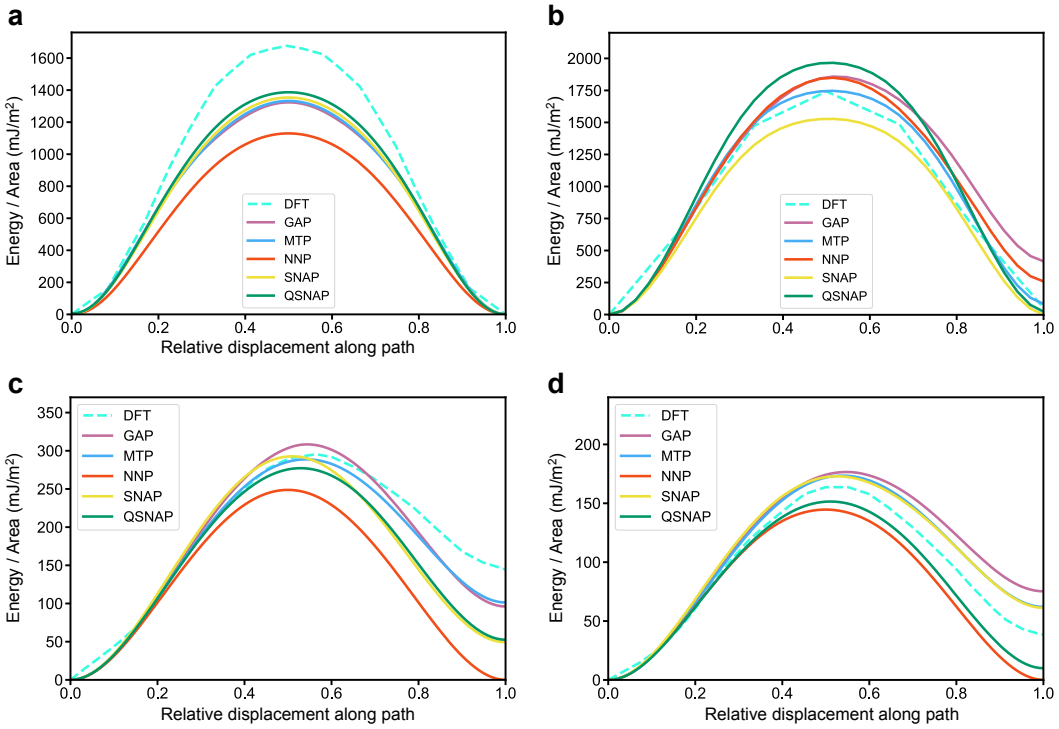



Figure 5: Cross sections of the relaxed $\gamma$ surfaces calculated with all four ML-IAPs with respect to DFT reference data.[57,86,95] (a) The relaxed (011) $\gamma$ surface along the $[\bar{1}\bar{1}1]$ direction for Mo. (b) The relaxed (111) glide plane along the [112] direction for Si. (c) The relaxed (111) $\gamma$ surface along the [112] direction for Ni. (d) The relaxed (111) $\gamma$ surface along the [112] direction for Cu.

Table 3: Calculated unstable stacking fault energy $\gamma_{us}$ of (011) $\gamma$ surface along the $[\bar{1}\bar{1}1]$ direction for Mo, (111) glide plane along the [112] direction for Si and (111) $\gamma$ surface along the [112] direction for Ni and Cu. Lowest absolute errors with respect to DFT are bolded for ease of reference. Error percentages with respect to DFT values are shown in parentheses.

| $\gamma_{us}$(mJ/m$^2$) | Mo (011) $[\bar{1}\bar{1}1]$ | Si (111) [112] | Ni (111) [112] | Cu (111) [112] |
|---|---|---|---|---|
| DFT | 1677[86] | 1740[57] | 289[95] | 164[95] |
| GAP | 1324 (−21.0%) | 1858 (6.8%) | 308 (6.6%) | 177 (7.9%) |
| MTP | 1333 (−20.5%) | **1747 (0.4%)** | **288 (-0.3%)** | 173 (5.5%) |
| NNP | 1130 (−32.6%) | 1849 (6.3%) | 248 (−14.2%) | 145 (−11.6%) |
| SNAP | 1354 (−19.3%) | 1528 (−12.2%) | 292 (1.0%) | **172 (4.9%)** |
| qSNAP | **1387 (-17.3%)** | 1965 (12.9%) | 277 (−4.2%) | 151 (−7.9%) |

## Accuracy in equations of state

To provide an evaluation of the performance of ML-IAPs far from equilibrium, we have computed a pairwise comparison of the equation of state (EOS) curves for all elements studied using the $\Delta_{\mathrm{EOS}}$ gauge of Lejaeghere et al.[96–98] The $\Delta_{\mathrm{EOS}}$ gauge, which has been used to evaluate accuracy differences between DFT codes, is the root-mean-square difference between two EOS curves over a ±6% interval around the equilibrium volume, defined as follows:

$$\Delta_{\mathrm{EOS}} = \sqrt{\frac{\int_{0.94V_0}^{1.06V_0} [E^{\mathrm{a}}(V) - E^{\mathrm{b}}(V)]^2 dV}{0.12V_0}} \tag{8}$$

where $E^{\mathrm{a}}$ and $E^{\mathrm{b}}$ denote energies computed using methods a and b, respectively.

Figure 6(b) shows the $\Delta_{\mathrm{EOS}}$ values of various machine learning models with respect to DFT reference data for different elemental systems as well as the EOS curves of these ML-IAPs. In all cases, the $\Delta_{\mathrm{EOS}}$ for all ML-IAPs for all elements are within 2 meV/atom, which is the threshold for "indistinguishable EOS" previously used in evaluating different DFT codes.[99] It is noteworthy that despite the relatively high prediction errors of SNAP models presented in Figure 3(a), they perform considerably better in predicting the EOS curves, with all the $\Delta_{\mathrm{EOS}}$ lower than 1 meV/atom across different chemistries. The NNP models

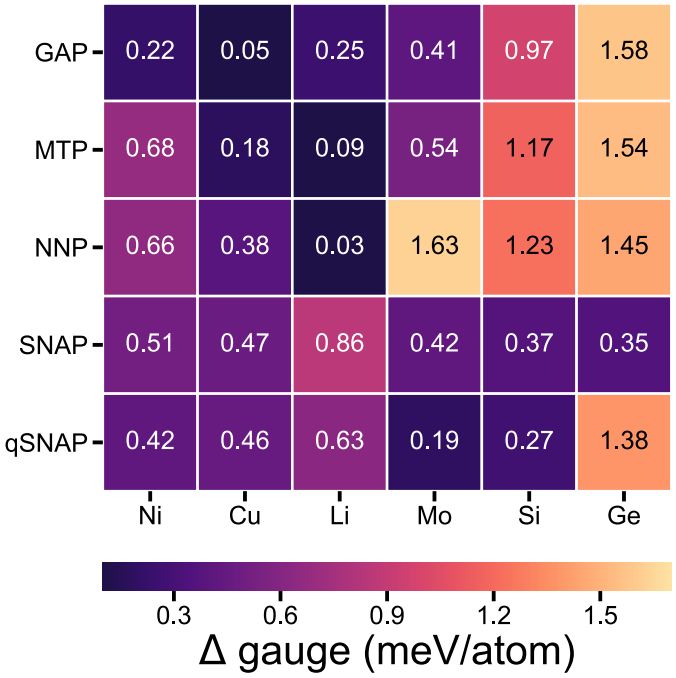
deviate slightly from DFT curves at both tensile and compressive strains for fcc systems, while for diamond systems, the deviation of the NNP models from DFT curve is comparable with those of GAP and MTP models, as evidenced in $\Delta$ gauge comparison. In general, it is more challenging to give highly accurate predictions of EOS in diamond system than in fcc and bcc systems. In addition to the DFT-level accuracy in equations of state prediction, the predicted phonon dispersion curves by all ML-IAPs investigated in this work are in excellent agreement with the DFT reference (see Figure S5 - S10 in the Supplementary Information).

## Accuracy in molecular dynamics (MD) trajectories

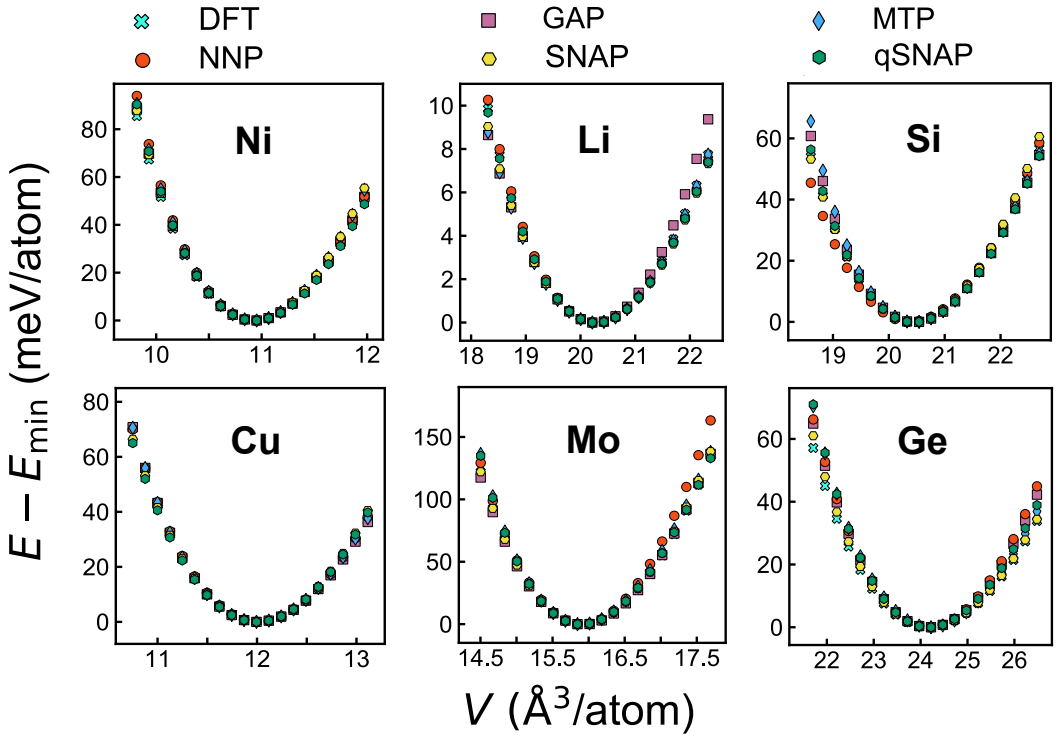
One of the principal applications of ML-IAPs is in molecular dynamics (MD) simulations. To assess the ability of the ML-IAPs to provide stable MD trajectories, we carried out $NVT$ MD simulations at 1,300 K (0.5× melting point) on a $3 \times 3 \times 3$ 54-atoms supercell of bulk Mo for 0.25 ns using LAMMPS with the different ML-IAPs. A total of 40 snapshots at an interval of 2.5 ps were sampled from each MD trajectory, and DFT static calculations were performed on these snapshots. Figure 7 shows the distribution of the errors in the energies and forces of sampled structures. In line with the previous results, the GAP and MTP models generally exhibit smaller errors in the energies and forces than the NNP, SNAP and qSNAP models. The GAP model has not only the lowest median but also the smallest interquartile range (IQR) in the errors in energies and forces. Somewhat interestingly, the NNP model has higher energy errors, but smaller force errors than SNAP and qSNAP. For consistency of comparison, all models shown here are the "optimal" models based on $\sim 100$ training structures. It is likely that a larger training set would improve the performance of the NNP and qSNAP models. (Figure 4).

## Accuracy in polymorphic energy differences

To evaluate the ability of the ML-IAPs to extrapolate to unseen data, we have computed the energy differences between the DFT ground state polymorph and a low-energy polymorph




(a) Δ gauge comparisons between ML-IAPs with respect to DFT




(b) Energy vs volume curves

Figure 6: Assessment of accuracy of ML-IAPs in predicting equation of state. (a) Δ gauge comparison provides quantitative estimate of deviation between the EOS curve from each ML-IAP with that of DFT. (b) EOS curves for all six elements using DFT and the four ML-IAPs.

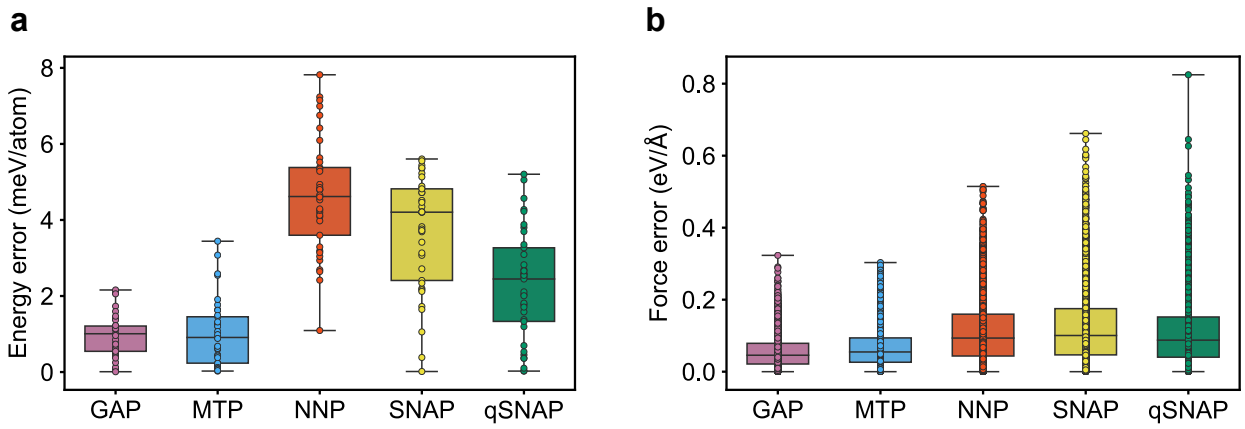



Figure 7: Error distributions in (a) predicted energies (b) predicted forces for sampled structures from MD simulations using each ML-IAP. The rectangular box indicates the interquartile range (IQR), while the line within the box indicates the median.

for each element, presented in Figure 8. The low-energy polymorphs correspond to the bcc, fcc, and wurtzite (hexagonal diamond) structures for the fcc, bcc, and diamond systems, respectively. It should be noted that only the ground state structures were used in training the ML-IAPs, and these low-energy polymorphs were not present in the training structures. Except for Li which has an extremely small energy difference between the fcc and bcc structures in DFT, all ML-IAPs are able to reproduce qualitatively the energy difference between polymorphs. For most systems, the ML-IAPs are able to reproduce energy differences between the polymorphs to within 10-20 meV/atom; the main exception is Mo, which exhibits a large energy difference between the fcc and bcc structures. One notable observation is that the GAP model shows the largest error in predicting the energy difference between the wurtzite and diamond structures in Si and Ge compared to the other ML-IAPs, despite having relatively low RMSE in predicted energies in these systems (see Figure 3(a)). We believe that this may be due to the fact that the active features in GAP diamond models did not cover the polymorph phase region and the Gaussian process is sensitive to unseen data, while the other IAPs are able to extrapolate the interactions to this unseen configuration more effectively. To test this hypothesis, we have constructed two additional Si GAP models derived by augmenting the original Si GAP model with (a) features close to the wurtzite

polymorph, and (b) features that are exactly correspond to the wurtzite polymorph. As seen from Figure S4, the predicted energy differences between the wurtzite and diamond polymorphs are much closer to the DFT values for the augmented Si GAP models. A principal component analysis further shows that the wurtzite polymorph lies in a feature space that is not covered by the original training data. Somewhat surprisingly, the linear SNAP model exhibits among the best performance in reproducing the polymorphic energy differences across all systems, outperforming even the GAP and MTP for Mo, Si and Ge, despite having substantially larger RMSEs in energies and forces.

# Conclusions

We have performed a comprehensive unbiased evaluation of the GAP, MTP, NNP, SNAP, and qSNAP ML-IAP models using consistently-generated DFT data on six elemental systems spanning different crystal structures (fcc, bcc, and diamond), chemistries (main group metals, transition metals and semiconductors) and bonding (metallic and covalent). This evaluation is carried out across three key metrics that are of critical importance for any potential user of these ML-IAPs:

1. Accuracy in predicted energies, forces and properties for both seen and unseen structures;

2. Training data requirements, which influence the number of expensive DFT computations that have to be performed to train an ML-IAP to a given accuracy; and

3. Computational cost, which influences the size of the systems on which computations can be performed for a given computing budget.

These three metrics are inextricably linked - for all the four ML-IAPs, an increase in number of DOF (with increase in computational cost) and increase in training structures generally leads to higher accuracy, though diminishing returns are observed beyond a certain

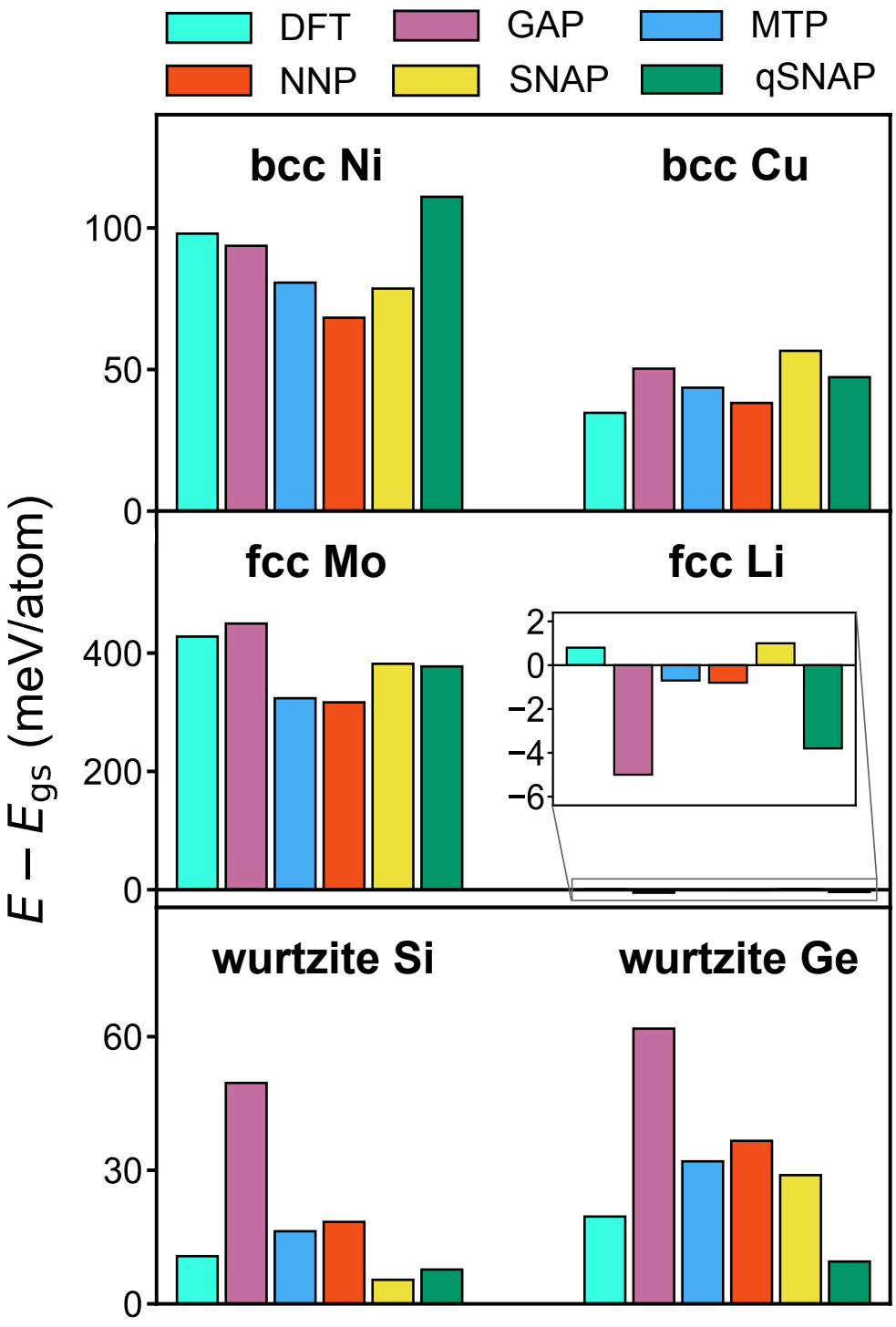



Figure 8: Calculated energetic differences between the typical low energy polymorph and ground-state polymorph of each elemental system. The inset shows the magnified bar chart for Li system due to its relatively small range. The typical low energy polymorph is indicated with the label above each bar chart.

number of DOFs. We demonstrate the application of the Pareto frontier as a means to identify the optimal trade-offs between these metrics. For all ML-IAPs, we find that there is an "optimal" configuration at which further expansion of the number of DOF yields little improvement in accuracy with increases in computational cost. We find that all ML-IAPs are able to achieve near-DFT accuracy in predicting energies, forces and material properties, substantially outperforming traditional IAPs. The GAP and MTP models exhibit the smallest RMSEs in energies and forces. However, the GAP models are among the most computationally expensive for a given accuracy (based on current implementations) and show poor extrapolability to higher energy polymorphs in the diamond systems. Indeed, the simple linear SNAP model, which has among the highest RMSEs in predicted energies and forces, show the best extrapolability to higher energy polymorphs as well as reproducing the equations of state for the diamond systems. The NNP and qSNAP models show relatively high RMSEs in energies with small data sizes, but these can be mitigated with increases in training data.

Another somewhat surprising conclusion is also that even with relatively small training data sets of ∼100-200 structures, the GAP, MTP and SNAP models appear to be reasonably well-converged to meV atom$^{-1}$ accuracy in energies and 0.01 eV Å$^{-1}$ accuracy in forces. The NNP and qSNAP models can be further improved with larger training data sets, but the RMSEs even at $\sim$ 100 structures are not excessively high. As training cost is a key performance metric, we have not attempted to further converge the NNP and qSNAP models beyond $\sim$ 800 structures (Figure 4), which is already several times the benchmarked training data set size. We attribute this performance to the training data generation procedure, which is aimed at sampling a diversity of structures from both ground state and multi-temperature AIMD simulations. In other words, the diversity of training data is arguably a more important consideration than quantity.

Finally, we will note several limitations of this study, which are possible avenues for future work. Firstly, no attempt was made include non-elemental (binary, ternaries, etc.)

systems in this study for practical considerations of computational cost in generating the large number of training structures needed and the diverse range of bonding types (e.g., metallic, ionic, covalent, and van der Waals, or a mixture of these). Second, we have chosen to focus on only ML-IAPs based on local environment descriptors. It should be noted that there are similarities between some of the descriptors. Both the SOAP and bispectrum (SNAP) descriptors are derived from the local atomic density function, while the NNP and MTP are based on parameterizations of the radial and angular distributions using various basis functions. There are alternative ML-IAPs that rely on global representations[100] that were not covered in this work. Finally, we have not attempted to combine the different local environment descriptors (symmetry functions, SOAP, bispectrum, moment tensors) with different ML frameworks (linear regression, Gaussian process regression, neural network, etc.). Instead, we rely on existing implementations of ML-IAPs as reported. The choice of descriptor affects how efficiently diverse local environments can be encoded, while the choice of ML framework determines the functional flexibility in mapping the relationship between descriptors and energies/forces. Ultimately, the performance of an ML-IAP is related to the atomic environment descriptor and the ML methodology. Our work has identified some differences in performances as being related to the choice of the ML model. For example, the quadratic qSNAP model can be converged to substantially higher accuracies than the linear SNAP model, albeit at increased risk of overfitting. The poorer extrapolability in the diamond systems and higher computational cost of the GAP can be traced to the use of Gaussian process regression. Constructing ML-IAPs using descriptor-model combinations can potentially yield further insights into the interplay between atomic descriptor and ML framework as well as better trade-offs between accuracy and cost for a particular application.

## Supporting Information

Detailed formalism and the optimized parameters and hyperparameters of all four ML-IAPs investigated; radial distribution function (RDF) of high temperature AIMD data; RMSEs of energy normalized to the ground state energy per atom for each system; learning curve of errors and forces in terms of size of training data; proof of concept of the poor performance of GAP model in transferability; phonon dispersion curve of ML-IAPs; trade-off between computational cost and test error of ML-IAPs with respect to Ni, Cu, Li, Si and Ge system; and linear scaling relationship of IAPs

## Acknowledgments

This work was primarily supported by the Office of Naval Research (ONR) Young Investigator Program (YIP) under Award No. N00014-16-1-2621. The authors also acknowledge computational resources provided by Triton Shared Computing Cluster (TSCC) at the University of California, San Diego, the National Energy Research Scientific Computing Center (NERSC), and the Extreme Science and Engineering Discovery Environment (XSEDE) supported by the National Science Foundation under Grant No. ACI-1053575. J. Behler thanks the Deutsche Forschungsgemeinschaft for a Heisenberg professorship (Be3264/11-2, project number 329898176). A. V. Shapeev was supported by the Russian Science Foundation under Grant No. 18-13-00479. A. Thompson and M. Wood are employees of Sandia National Laboratories, a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

# References


(1) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(2) Sham, L. J.; Schlüter, M. Density-Functional Theory of the Energy Gap. *Phys. Rev. Lett.* **1983**, *51*, 1888–1891.

(3) Zhang, I. Y.; Xu, X.; Jung, Y.; Goddard, W. A. A Fast Doubly Hybrid Density Functional Method Close to Chemical Accuracy Using a Local Opposite Spin Ansatz. *Proc. Natl. Acad. Sci.* **2011**, *108*, 19896–19900.

(4) Zhang, Y.; Xu, X.; Goddard, W. A. Doubly Hybrid Density Functional for Accurate Descriptions of Nonbond Interactions, Thermochemistry, and Thermochemical Kinetics. *Proc. Natl. Acad. Sci.* **2009**, *106*, 4963–4968.

(5) Ji, H.; Shao, Y.; Goddard, W. A.; Jung, Y. Analytic Derivatives of Quartic-Scaling Doubly Hybrid XYGJ-OS Functional: Theory, Implementation, and Benchmark Comparison with M06-2X and MP2 Geometries for Nonbonded Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 1971–1976.

(6) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(7) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897–8909.

(8) Van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.

(9) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901.

(10) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87*, 184115.

(11) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(12) Behler, J. Neural Network Potential-Energy Surfaces in Chemistry: A Tool for Large-Scale Simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930.

(13) Szlachta, W. J.; Bartók, A. P.; Csányi, G. Accuracy and Transferability of Gaussian Approximation Potential Models for Tungsten. *Phys. Rev. B* **2014**, *90*, 104108.

(14) Dragoni, D.; Daff, T. D.; Csányi, G.; Marzari, N. Achieving DFT Accuracy with a Machine-Learning Interatomic Potential: Thermomechanics and Defects in Bcc Ferromagnetic Iron. *Phys. Rev. Mater.* **2018**, *2*, 013808.

(15) Thompson, A.; Swiler, L.; Trott, C.; Foiles, S.; Tucker, G. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **2015**, 15.

(16) Chen, C.; Deng, Z.; Tran, R.; Tang, H.; Chu, I.-H.; Ong, S. P. Accurate Force Field for Molybdenum by Machine Learning Large Materials Data. *Phys. Rev. Mater.* **2017**, *1*, 043603.

(17) Li, X.-G.; Hu, C.; Chen, C.; Deng, Z.; Luo, J.; Ong, S. P. Quantum-Accurate Spectral Neighbor Analysis Potential Models for Ni-Mo Binary Alloys and Fcc Metals. *Phys. Rev. B* **2018**, *98*, 094104.

(18) Deng, Z.; Chen, C.; Li, X.-G.; Ong, S. P. An Electrostatic Spectral Neighbor Analysis Potential for Lithium Nitride. *npj Comput. Mater.* **2019**, *5*, 75.

(19) Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul.* **2016**, *14*, 1153–1173.

(20) Podryabinkin, E. V.; Shapeev, A. V. Active Learning of Linearly Parametrized Interatomic Potentials. *Comput. Mater. Sci.* **2017**, *140*, 171–180.

(21) Gubaev, K.; Podryabinkin, E. V.; Hart, G. L. W.; Shapeev, A. V. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials. *Comput. Mater. Sci.* **2019**, *156*, 148–156.

(22) Botu, V.; Ramprasad, R. Adaptive Machine Learning Framework to Accelerate *Ab Initio* Molecular Dynamics. *Int. J. Quantum Chem.* **2015**, *115*, 1074–1083.

(23) Botu, V.; Ramprasad, R. Learning Scheme to Predict Atomic Forces and Accelerate Materials Simulations. *Phys. Rev. B* **2015**, *92*, 094306.

(24) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2017**, *121*, 511–522.

(25) Li, Z.; Kermode, J. R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*, 096405.

(26) Schütt, K. T.; Kessel, P.; Gastegger, M.; Nicoli, K. A.; Tkatchenko, A.; Müller, K.-R. SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput.* **2019**, *15*, 448–455.

(27) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham Equations with Machine Learning. *Nat. Commun.* **2017**, *8*, 872.

(28) Rupp, M. Machine Learning for Quantum Mechanics in a Nutshell. *Int. J. Quantum Chem.* **2015**, *115*, 1058–1073.

(29) Huang, Y.; Kang, J.; Goddard, W. A.; Wang, L.-W. Density Functional Theory Based Neural Network Force Fields from Energy Decompositions. *Phys. Rev. B* **2019**, *99*, 064103.

(30) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326–2331.

(31) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(32) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.

(33) Bereau, T.; DiStasio, R. A.; Tkatchenko, A.; von Lilienfeld, O. A. Non-Covalent Interactions across Organic and Biological Subsets of Chemical Space: Physics-Based Potentials Parametrized from Machine Learning. *J. Chem. Phys.* **2018**, *148*, 241706.

(34) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192–3203.

(35) Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 Model Chemistry: A Neural Network Augmented with Long-Range Physics. *Chem. Sci.* **2018**, *9*, 2261–2269.

(36) Nguyen, T. T.; Székely, E.; Imbalzano, G.; Behler, J.; Csányi, G.; Ceriotti, M.; Götz, A. W.; Paesani, F. Comparison of Permutationally Invariant Polynomials, Neural Networks, and Gaussian Approximation Potentials in Representing Water Interactions through Many-Body Expansions. *J. Chem. Phys.* **2018**, *148*, 241725.

(37) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Canova, F. F.; Ranawat, Y. S.;

Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of Descriptors for Machine Learning in Materials Science. *arXiv:1904.08875* **2019**,

(38) Nyshadham, C.; Rupp, M.; Bekker, B.; Shapeev, A. V.; Mueller, T.; Rosenbrock, C. W.; Csányi, G.; Wingate, D. W.; Hart, G. L. W. Machine-Learned Multi-System Surrogate Models for Materials Prediction. *npj Comput. Mater.* **2019**, *5*, 51.

(39) Balamane, H.; Halicioglu, T.; Tiller, W. A. Comparative study of silicon empirical interatomic potentials. *Phys. Rev. B* **1992**, *46*, 2250–2279.

(40) O'connor, D.; Biersack, J. Comparison of theoretical and empirical interatomic potentials. *Nucl. Instrum. Methods Phys. Res., Sect. B* **1986**, *15*, 14–19.

(41) Godet, J.; Pizzagalli, L.; Brochard, S.; Beauchamp, P. Comparison between classical potentials and ab initio methods for silicon under large shear. *J. Phys.: Condens. Matter* **2003**, *15*, 6943.

(42) Behler, J. Constructing High-Dimensional Neural Network Potentials: A Tutorial Review. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050.

(43) Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828–12840.

(44) Mocanu, F. C.; Konstantinou, K.; Lee, T. H.; Bernstein, N.; Deringer, V. L.; Csányi, G.; Elliott, S. R. Modeling the Phase-Change Memory Material, $Ge_2Sb_2Te_5$, with a Machine-Learned Interatomic Potential. *J. Phys. Chem. B* **2018**, *122*, 8998–9006.

(45) Wood, M. A.; Cusentino, M. A.; Wirth, B. D.; Thompson, A. P. Data-Driven Material Models for Atomistic Simulation. *Phys. Rev. B* **2019**, *99*, 184305.

(46) Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2011**, *134*, 074106.

(47) Artrith, N.; Urban, A. An Implementation of Artificial Neural-Network Potentials for Atomistic Materials Simulations: Performance for TiO2. *Comput. Mater. Sci.* **2016**, *114*, 135–150.

(48) Morawietz, T.; Singraber, A.; Dellago, C.; Behler, J. How van der Waals interactions determine the unique properties of water. *Proc. Natl. Acad. Sci.* **2016**, *113*, 8368–8373.

(49) Quaranta, V.; Hellström, M.; Behler, J. Proton-transfer mechanisms at the water–ZnO interface: the role of presolvation. *J. Phys. Chem. Lett.* **2017**, *8*, 1476–1483.

(50) Eckhoff, M.; Behler, J. From Molecular Fragments to the Bulk: Development of a Neural Network Potential for MOF-5. *J. Chem. Theory Comput.* **2019**,

(51) Artrith, N.; Morawietz, T.; Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **2011**, *83*, 153101.

(52) Li, W.; Ando, Y.; Minamitani, E.; Watanabe, S. Study of Li Atom Diffusion in Amorphous Li3PO4 with Neural Network Potential. *J. Chem. Phys.* **2017**, *147*, 214106.

(53) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(54) Deringer, V. L.; Pickard, C. J.; Csányi, G. Data-Driven Learning of Total and Local Energies in Elemental Boron. *Phys. Rev. Lett.* **2018**, *120*, 156001.

(55) Deringer, V. L.; Csányi, G. Machine Learning Based Interatomic Potential for Amorphous Carbon. *Phys. Rev. B* **2017**, *95*, 094203.

(56) Rowe, P.; Csányi, G.; Alfè, D.; Michaelides, A. Development of a Machine Learning Potential for Graphene. *Phys. Rev. B* **2018**, *97*, 054303.

(57) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine Learning a General-Purpose Interatomic Potential for Silicon. *Phys. Rev. X* **2018**, *8*, 041048.

(58) Deringer, V. L.; Bernstein, N.; Bartók, A. P.; Cliffe, M. J.; Kerber, R. N.; Marbella, L. E.; Grey, C. P.; Elliott, S. R.; Csányi, G. Realistic Atomistic Structure of Amorphous Silicon from Machine-Learning-Driven Molecular Dynamics. *J. Phys. Chem. Lett.* **2018**, *9*, 2879–2885.

(59) Wood, M. A.; Thompson, A. P. Extending the Accuracy of the SNAP Interatomic Potential Form. *J. Chem. Phys.* **2018**, *148*, 241721.

(60) Novoselov, I.; Yanilkin, A.; Shapeev, A.; Podryabinkin, E. Moment Tensor Potentials as a Promising Tool to Study Diffusion Processes. *Comput. Mater. Sci.* **2019**, *164*, 46–56.

(61) Podryabinkin, E. V.; Tikhonov, E. V.; Shapeev, A. V.; Oganov, A. R. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Phys. Rev. B* **2019**, *99*, 064114.

(62) Novikov, I. S.; Suleimanov, Y. V.; Shapeev, A. V. Automated calculation of thermal rate coefficients using ring polymer molecular dynamics and machine-learning interatomic potentials with active learning. *Phys. Chem. Chem. Phys.* **2018**, *20*, 29503–29512.

(63) de Jong, M.; Chen, W.; Angsten, T.; Jain, A.; Notestine, R.; Gamst, A.; Sluiter, M.; Krishna Ande, C.; van der Zwaag, S.; Plata, J. J.; Toher, C.; Curtarolo, S.; Ceder, G.; Persson, K. A.; Asta, M. Charting the Complete Elastic Properties of Inorganic Crystalline Compounds. *Sci. Data* **2015**, *2*, 150009.

(64) Tran, R.; Xu, Z.; Radhakrishnan, B.; Winston, D.; Sun, W.; Persson, K. A.; Ong, S. P. Data Descriptor: Surface Energies of Elemental Crystals. *Sci. Data* **2016**, *23*, 1–13.

(65) An Exploration of the Grain Boundaries, Surfaces, and Wulff shapes of the Elements. http://crystalium.materialsvirtuallab.org/ (accessed 2016).

(66) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for *Ab Initio* Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(67) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953–17979.

(68) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(69) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319.

(70) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G.-M.; Hautier, G.; Gunter, D.; Persson, K. A. FireWorks: A Dynamic Workflow System Designed for High-Throughput Applications: FireWorks: A Dynamic Workflow System Designed for High-Throughput Applications. *Concurrency Computat.: Pract. Exp.* **2015**, *27*, 5037–5059.

(71) Lee, B.-J.; Shim, J.-H.; Baskes, M. I. Semiempirical Atomic Potentials for the Fcc Metals Cu, Ag, Au, Ni, Pd, Pt, Al, and Pb Based on First and Second Nearest-Neighbor Modified Embedded Atom Method. *Phys. Rev. B* **2003**, *68*, 144112.

(72) Foiles, S. M.; Baskes, M. I.; Daw, M. S. Embedded-Atom-Method Functions for the Fcc Metals Cu, Ag, Au, Ni, Pd, Pt, and Their Alloys. *Phys. Rev. B* **1986**, *33*, 7983–7991.

(73) Song, Y.; Yang, R.; Li, D.; Wu, W. T.; Guo, Z. X. Calculation of Theoretical Strengths and Bulk Moduli of Bcc Metals. *Phys. Rev. B* **1999**, *59*, 14220–14225.

(74) Adams, J. B.; Foiles, S. M. Development of an Embedded-Atom Potential for a Bcc Metal: Vanadium. *Phys. Rev. B* **1990**, *41*, 3316–3328.

(75) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing Molecules and Solids across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.

(76) Baskes, M. I.; Nelson, J. S.; Wright, A. F. Semiempirical Modified Embedded-Atom Potentials for Silicon and Germanium. *Phys. Rev. B* **1989**, *40*, 6085–6100.

(77) Singraber, A.; Behler, J.; Dellago, C. Library-Based *LAMMPS* Implementation of High-Dimensional Neural Network Potentials. *J. Chem. Theory Comput.* **2019**, *15*, 1827–1840.

(78) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1 – 19.

(79) Zhang, Y.; Hu, C.; Jiang, B. Embedded Atom Neural Network Potentials: Efficient and Accurate Machine Learning with a Physically Inspired Representation. *J. Phys. Chem. Lett.* **2019**, *10*, 4962–4967.

(80) Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I.; Wojna, Z.; Song, Y.; Guadarrama, S.; Murphy, K. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* **2017**, 3296–3297.

(81) Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R. End-to-End Symmetry Preserving Inter-Atomic Potential Energy Model for Finite and Extended Systems. *Proceedings of the 32nd International Conference on Neural Information Processing Systems; Curran Associates Inc.: Montreal, Canada* **2018**, 4441–4451.

(82) Cheng, B.; Engel, E. A.; Behler, J.; Dellago, C.; Ceriotti, M. Ab Initio Thermodynamics of Liquid and Solid Water. *Proc. Natl. Acad. Sci.* **2019**, *116*, 1110–1115.

(83) Zhou, X. W.; Johnson, R. A.; Wadley, H. N. G. Misfit-Energy-Increasing Dislocations in Vapor-Deposited CoFe/NiFe Multilayers. *Phys. Rev. B* **2004**, *69*, 144113.

(84) Nichol, A.; Ackland, G. J. Property trends in simple metals: An empirical potential approach. *Phys. Rev. B* **2016**, *93*, 184101.

(85) Asadi, E.; Asle Zaeem, M.; Nouranian, S.; Baskes, M. I. Two-Phase Solid–Liquid Coexistence of Ni, Cu, and Al by Molecular Dynamics Simulations Using the Modified Embedded-Atom Method. *Acta Mater.* **2015**, *86*, 169–181.

(86) Park, H.; Fellinger, M. R.; Lenosky, T. J.; Tipton, W. W.; Trinkle, D. R.; Rudin, S. P.; Woodward, C.; Wilkins, J. W.; Hennig, R. G. *Ab Initio* Based Empirical Potential Used to Study the Mechanical Properties of Molybdenum. *Phys. Rev. B* **2012**, *85*, 214121.

(87) Lenosky, T. J.; Sadigh, B.; Alonso, E.; Bulatov, V. V.; de la Rubia, T. D.; Kim, J.; Voter, A. F.; Kress, J. D. Highly optimized empirical potential model of silicon. *Modell. Simul. Mater. Sci. Eng.* **2000**, *8*, 825–841.

(88) Kumagai, T.; Izumi, S.; Hara, S.; Sakai, S. Development of bond-order potentials that can reproduce the elastic constants and melting point of silicon for classical molecular dynamics simulation. *Comput. Mater. Sci.* **2007**, *39*, 457 – 464.

(89) Mahdizadeh, S. J.; Akhlamadi, G. Optimized Tersoff empirical potential for germanene. *J. Mol. Graphics Modell.* **2017**, *72*, 1 – 5.

(90) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978–9985.

(91) Hill, R. The Elastic Behaviour of a Crystalline Aggregate. *Proc. Phys. Soc., London, Sect. A* **1952**, *65*, 349–354.

(92) Frederiksen, S. L.; Jacobsen, K. W. Density Functional Theory Studies of Screw Dislocation Core Structures in Bcc Metals. *Philos. Mag.* **2003**, *83*, 365–375.

(93) Juan, Y.-M.; Kaxiras, E. Generalized Stacking Fault Energy Surfaces and Dislocation Properties of Silicon: A First-Principles Theoretical Study. *Philos. Mag. A* **1996**, *74*, 1367–1384.

(94) Zimmerman, J. A.; Gao, H.; Abraham, F. F. Generalized Stacking Fault Energies for Embedded Atom FCC Metals. *Modell. Simul. Mater. Sci. Eng.* **2000**, *8*, 103–115.

(95) Hunter, A.; Zhang, R. F.; Beyerlein, I. J. The Core Structure of Dislocations and Their Relationship to the Material $\gamma$-Surface. *J. Appl. Phys.* **2014**, *115*, 134314.

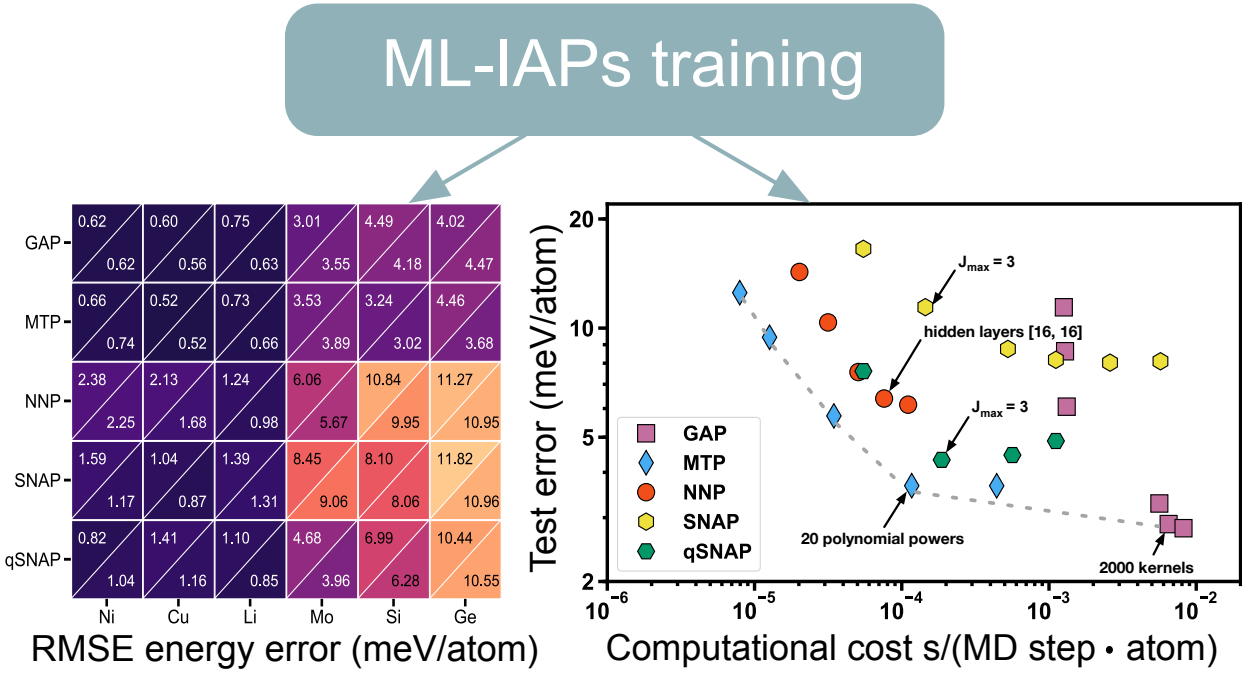
(96) Lejaeghere, K.; Van Speybroeck, V.; Van Oost, G.; Cottenier, S. Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals. *Crit. Rev. Solid State Mater. Sci.* **2014**, *39*, 1–24.

(97) Csonka, G. I.; Perdew, J. P.; Ruzsinszky, A.; Philipsen, P. H. T.; Lebègue, S.; Paier, J.; Vydrov, O. A.; Ángyán, J. G. Assessing the Performance of Recent Density Functionals for Bulk Solids. *Phys. Rev. B* **2009**, *79*, 155107.

(98) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Tests of a Ladder of Density Functionals for Bulk Solids and Surfaces. *Phys. Rev. B* **2004**, *69*, 075102.

(99) Lejaeghere, K.; Bihlmayer, G.; Bjorkman, T.; Blaha, P.; Blugel, S.; Blum, V.; Caliste, D.; Castelli, I. E.; Clark, S. J.; Dal Corso, A.; et al., Reproducibility in Density Functional Theory Calculations of Solids. *Science* **2016**, *351*, 1–11.

(100) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3*, e1603015.


ML-IAPs training
GAP
MTP
NNP
SNAP
qSNAP
Ni
Cu
Li
Mo
Si
Ge
RMSE energy error (meV/atom)
Test error (meV/atom)
Computational cost s/(MD step • atom)
GAP
MTP
NNP
SNAP
qSNAP
hidden layers [16, 16]
20 polynomial powers
2000 kernels